Improving Image Quality of Single Plane Wave Ultrasound via Deep Learning Based Channel Compounding

Sven Rothlübbers*, Hannah Strohm*, Klaus Eickel†, Jürgen Jenne*,
Vincent Kuhlen*†, David Sinden* and Matthias Günther*†‡
Email: {sven.rothluebbers, hannah.strohm}@mevis.fraunhofer.de,
*Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany
†University of Bremen, Bremen, Germany
‡mediri GmbH, Heidelberg, Germany

***Abstract*—The emergence of data driven approaches such as Deep Learning has led to novel application of various aspects of science and engineering. It has recently entered the field of ultrasound image beamforming. In this work we investigate neural networks tailored to create images of the quality of multiple compounded plane wave excitations from the data of the central angle (0°) excitation only. The proposed network is used to produce pixel-wise weights to weigh a standard delay-and-sum image from all channel data available to a pixel. It is found to produce higher quality images than the classical reference reconstruction from the 0° angle data.**

***Index Terms*—Ultrasound, Deep Learning, Beamforming, Plane Wave Imaging, Pixel-Weighting**

I. Introduction

Planewave imaging can be performed at ultrafast frame rates, offering the potential for new imaging capabilities, such as detailed three-dimension cardiac function through ultrafast power Doppler [1]. In order to image at ultrafast frame rates, only a small number of plane waves can be acquired. However, when using only a few plane waves per image, the quality of the resulting image deteriorates, reducing the diagnostic capabilities. There is a clear need for fast reconstruction algorithms which operates on only few acquisitions, addressed here as a contribution to task 1 of the CUBDL-challenge [2], [3].

II. Background

Many ultrasound (US) plane wave reconstruction algorithms in the spatio-temporal domain consist of computing receive delays for each raw data channel towards a spatial point of interest and then summing the raw data. For basic delay-and-sum imaging, a single plane wave excitation yields N_C values (channel data) that when summed with equal weights can yield a low-quality image. In order to improve image quality, algorithms have been proposed that compute weights for the channel data (channel weights) or the summed pixel data (pixel weights), such as for example the computationally expensive minimum variance (MV) beamforming [4] and sign coherence factor (SCF) imaging [5], respectively.

This work was supported by Fraunhofer Funding Discover 600725 Ultra-Deep and Prepare 601100 Theranus, which is gratefully acknowledged.

The acquisition of N_A different angles allows compounding multiple single shot images into a higher quality image. In general, the order in which N_C channels and N_A shots are compounded may be varied [6] and in some cases the compounding of $N_C \times N_A$ values can even be performed in a joint expression. In united sign coherence factor (USCF) imaging [7], the pixel weight is computed as

$$w_{\mathrm{USCF}} = \left| 1 - \sqrt{1 - \left(\frac{1}{N_C N_A} \sum_{i=1}^{N_C} \sum_{j=1}^{N_A} b_{i,j} \right)^2} \right|^p$$

where $b_{i,j}$ is the sign of the current pixels raw data element of channel i and acquisition j and p can be used as a global scaling power ($p = 1$ by default).

In this work, a neural network is used to replace a weighting function in the beamforming pipeline with the goal of producing high quality images (multi-angle USCF) from a single shot at angle 0°.

III. Materials & Methods

A. Data

For training, 107 US raw data sets of a phantom (Model 054GS, CIRS, Norfolk, VA) were acquired with a 128-element linear transducer (DiPhAS, Fraunhofer IBMT, Sankt Ingbert, Germany) operating at 4 MHz. High-quality target images were reconstructed using a modified (with $p = 0.9$ instead of $p = 1$) multi-angle USCF, utilizing data from seven plane wave angles. The reconstruction grid was chosen with an equidistant isotropic pixel spacing of a third of the wavelength and positioned such that artifact prone areas such as for example near the transducer were excluded.

The publicly available PICMUS dataset [8] was used to test the model.

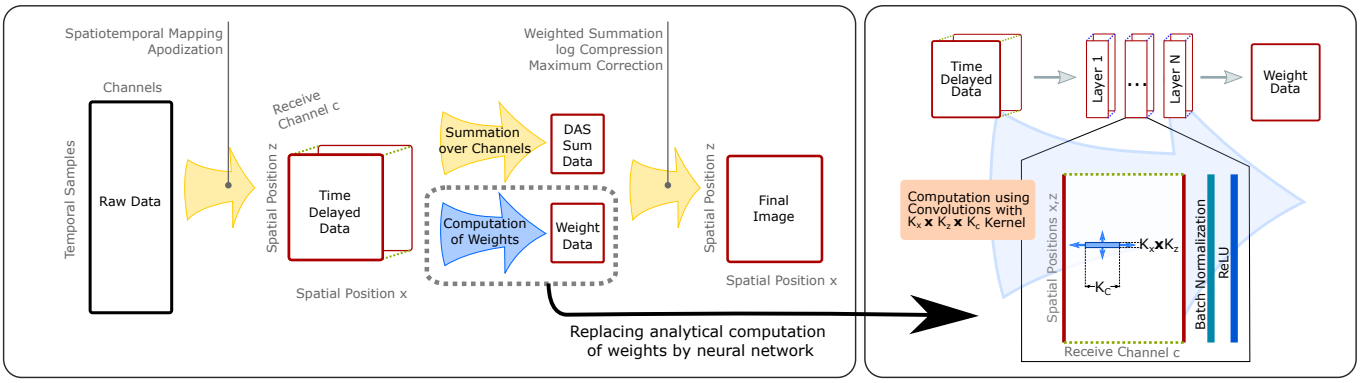

Fig. 1. Integration of the network into the beamforming pipeline. Within a weighted summation delay-and-sum pipeline the algorithm determining pixel (or channel) weights from channel data is replaced by a neural network (left). The examined networks apply convolutional layers that generally allow for convolutions in the x ,z and c domain, followed by batch normalization and ReLU activation (right).

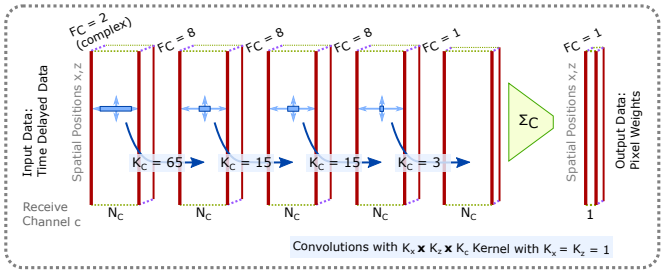

Fig. 2. The proposed final network architecture applies four layers with 1D convolutions ($K_x = K_z = 1$) along the channel domain before summing up the signal into the final pixel weights.

B. General Beamforming Pipeline

The general beamforming pipeline, illustrated in Fig. 1, consists of multiple stages. The input data is assumed to be complex valued and if not, extended by an imaginary component via a Hilbert transform. Preprocessing steps such as spatio-temporal mapping (delay computation) and apodization are followed by the central stage for channel or pixel weight computation. After weighting, the data is post-processed via log compression and normalization by its maximum. In this work, the computation of weights is replaced by a neural network.

C. Networks

Different network architectures were explored to compute weighting factors for the central processing step in the beamforming pipeline, illustrated in Fig. 1. The networks take in time delayed, magnitude normalized, but not yet channel-summed complex valued data from the central angle, in order to create either channel or pixel weights. The regular beamforming pipeline then continues to post-process the data, before the result is compared to the target. The networks were trained to bring the resulting output closer towards the USCF target data.

Several architecture variations were investigated experimentally. All of them were fully convolutional, applying $K_x \times K_z \times K_c$ kernels with spatial dimensions x, z and channel dimension c to the time-delayed data (Fig. 1, right). Each convolution is proceeded by batch normalization and ReLU activation. The filter sizes as well as the number of convolutional layers were varied among approaches. Each filter produces individual F_C feature channels, with the number of network input channels always being $F_C = 2$ (real and imaginary part).

For 1D networks with $K_x = K_z = 1$, the convolution kernels operate along the US channel axis (Fig. 2), whereas for 2D and 3D the kernels may also span along the spatial xz dimensions $(K_x, K_z > 1)$ taking the US channel data from neighboring pixels into account. Two versions of the final compounding function were explored: In 'pixel weighting' a single weight is computed that weights the full channel sum. In 'channel weighting', each of the N_C channels is weighted by an own weight value during the summation.

To account for memory limitations during training and inference a patch-based approach was used, dividing input and target data into patches of size e.g. 200×200. Zero-padding was applied such that the size of the patches was preserved throughout the convolutions. The loss was computed as a linear combination of mean-squared error (MSE) and multiscale structural similarity (MS-SSIM) [9] loss on the log compressed, normalized final images as

$$\mathcal{L} = \varepsilon L_{\mathrm{MSE}} + L_{\mathrm{MS-SSIM}}.$$

A factor of $\varepsilon = 10^{-4}$ is chosen to bring the value ranges of the two loss components to a similar level and emphasize the structural similarity between network prediction and target image.

The best model for each training run was chosen according to the validation loss. Networks were implemented using PyTorch [10].

IV. Results

Final performance was evaluated by two experts subjectively judging results on a test set.

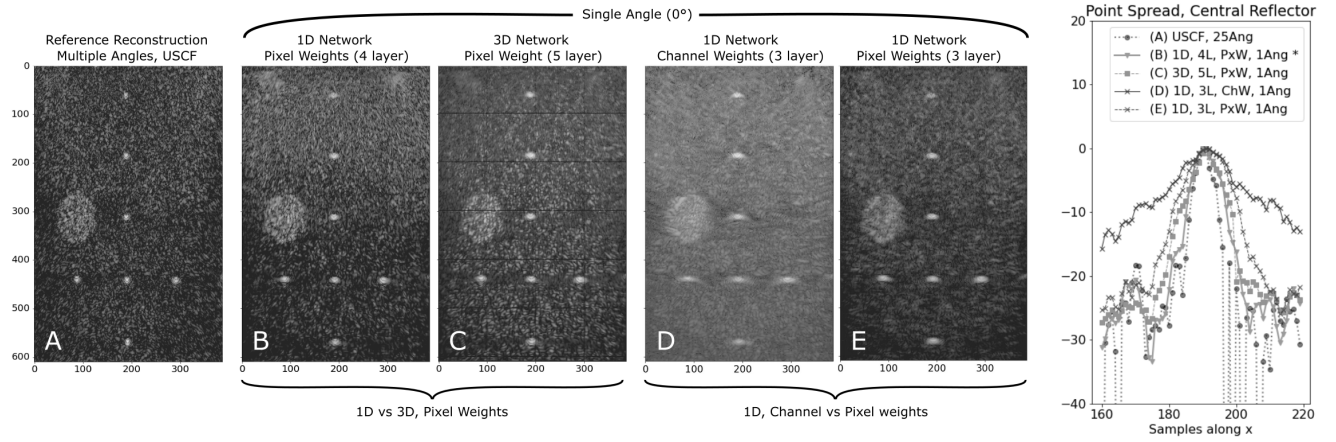

Fig. 3. Comparison of results for different architectures from the 0° angle (B-E) to the multi-angle reference (A) of unseen test data from the PICMUS [8] dataset. Subfigures B, C compare the proposed 1D pixel weighting network to a 3D version. In C the visible black lines (squares) of the 3D network output are an artifacts from spatial padding of limited size patches. Subfigures D, E compare a channel weighting to a pixel weighting implementation of three layers each. The right hand side subfigure depicts the lateral point spread for the central small scatterer at (190,440) for all five cases (curves averaged over 11 axial samples and maxima shifted to 0 dB).

Fig. 3 shows network predictions of four different architectures for a test image from the PICMUS dataset (subfigures B-E) compared to the reference reconstruction with USCF (subfigure A). The visual impression is well supported by the analysis of the lateral point spread for the discussed cases on the right hand side.

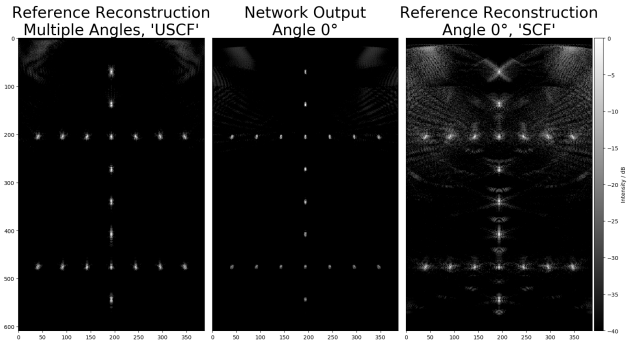

Fig. 4. Comparison of network output to multi and single angle reference reconstructions on unseen artificial test data from PICMUS [8] dataset.

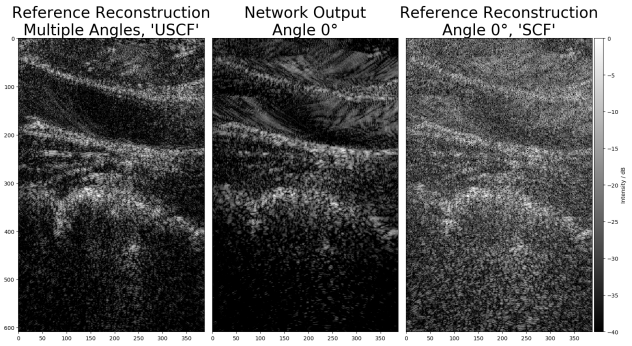

Fig. 5. Comparison of network output to multi and single angle reference reconstructions on unseen *in-vivo* test data from PICMUS [8] dataset.

Within the explored architectures training channel weights did not show a benefit over training pixel weights. This can be seen in subfigures (E, D), where training only a single weight per pixel (E) leads to better contrast compared to training individual channel weights (D). In the presented case, the lower scatterers for the channel weighting result appear more blurred, which can also clearly be seen in the increase point spread.

The choice of including spatial dimensions into the kernels, i.e. either using two (for example xc, zc) or three (xzc) dimension was also not shown to have a clearly favorable impact to the result in this work (subfigures B, C). The depicted point spread function even indicates a slight advantage of the 1D implementation. At the same time the 3D networks introduce a higher complexity, longer training and computation time. Also, padding in the xz domain can cause artifacts at patch borders if padding is not chosen appropriately (e.g. the applied zero filling leads to dark patch edges). This should be avoidable by overlapping patch areas, however.

Thus, the simple 1D network (B) with 4 layers is finally proposed (Fig. 2) to compute pixel-wise weighting to improve the quality of a single shot reconstruction. Kernel sizes (65,15,15,3) and feature map sizes (8,8,8,1) are used. Generally, using a large kernel in the first layer (i.e. $K_C = 65$) and a network depth of at least four layers produced reasonable results.

A comparison of the network reconstruction with the single shot SCF-beamformed reference image can be seen in Figs. 4 and 5. On phantom data, as Fig. 4 shows, a clear improvement in image sharpness and a reduction of artifacts can be observed. Generally, a subjective increase in contrast

and detail is found. However, in the PICMUS *in-vivo* images (e.g. Fig. 5) occasional streaking artifacts within angle 0° are not suppressed sufficiently. Generally, the network output is closer to the multi-angle reference than the single shot image is, yet not able to fully reproduce the multi-angle reference.

V. Discussion & Conclusion

This work compares experimentally different variants of network architectures to compute weighting factors for the compounding step in ultrasound beamforming.

Within the explored architectures, a decent performance can already be observed by networks of low complexity. Specifically, training pixel weighting outperforms training channel weighting. Furthermore, 2D and 3D approaches did not seem to be of advantage within the experiments. The motivation to include spatial dimensions was to allow the network to detect spatial patterns within the channel data. However, due to memory and training time constraints these networks were left with only relatively few layers (less than 10), which might hamper the detection of patterns. It remains an open question whether deeper architectures might be able to outperform the 1D version.

The proposed architecture produces higher quality images from a single plane wave excitation than the single angle reference reconstruction. There are a number of advantages of this architecture. Firstly, in contrast to fully-connected networks [11], the proposed fully-convolutional architecture allows for both the handling of images of varying sizes and ultrasound channels, allowing for data from different transducers.

Furthermore, as the network performs the reconstruction after applying time-delays, it can be integrated into other reconstruction pipelines. With fewer layers and feature maps (less than 8), the architecture has a lower network complexity than existing deep learning models [12]. The achieved reduction can be seen as an indication that the complexity of the reconstruction can generally be further reduced. It can be suspected that even more simple functional components might produce good results and also drive the architecture further towards an understanding of its inner functionality.

The use of a large first layer convolutional kernel indicates that it is necessary to provide the network with a large receptive field, which could reveal more information about the network functionality.

In this work the USCF algorithm was used as reference. Unlike minimum variance contrast, the USCF is computationally inexpensive. Thus, a computational advantage is not an argument for the proposed method, but that higher quality of images can be produced from a single frame.

Future extensions of this work could include an application to other reference algorithms, such as MV beamformed images. The integration of further image quality metrics, such as contrast-to-noise is favorable, even if only applied for validation purposes. Further, an application that uses few angles instead of a single one will have interesting application, especially in accelerating 3D ultrasound imaging.

References

[1] M. Correia *et al.*, "Quantitative imaging of coronary flows using 3D ultrafast Doppler coronary angiography," *Phys. Med. Biol.*, vol. 65, no. 10, p. 105013, 2020.

[2] "Challenge on ultrasound beamforming with deep learning (CUBDL)," 2019. [Online]. Available: http://dx.doi.org/10.21227/f0hn-8f92

[3] M. A. L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. J. G. van Sloun, and M. Mischi, "Challenge on ultrasound beamforming with deep learning (CUBDL)," in *2020 Proc. IEEE Int. Ultrason. Symp. (IUS)*. IEEE, 2020.

[4] J. Li and P. Stoica, *Robust Adaptive Beamforming*. Hoboken, NJ: Wiley, 2006.

[5] J. Camacho, M. Parrilla, and C. Fritsch, "Phase coherence imaging," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 56, no. 5, pp. 958–974, 2009.

[6] N. Q. Nguyen and R. W. Prager, "A spatial coherence approach to minimum variance beamforming for plane-wave compounding," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 65, no. 4, pp. 522–534, 2018.

[7] C. Yang, Y. Jiao, T. Jiang, Y. Xu, and Y. Cui, "A united sign coherence factor beamformer for coherent plane-wave compounding with improved contrast," *Appl. Sci.*, vol. 10, no. 7, p. 2250, 2020.

[8] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *2016 Proc. IEEE Int. Ultrason. Symp. (IUS)*. IEEE, 2016, pp. 1–4.

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc 37th Asilomar Conf. Signals, Systems & Computers*, vol. 2. IEEE, 2003, pp. 1398–1402.

[10] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., 2019, pp. 8024–8035.

[11] B. Luijten, R. Cohen, F. J. de Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. van Sloun, "Deep learning for fast adaptive beamforming," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*. IEEE, 2019, pp. 1333–1337.

[12] S. Khan, J. Huh, and J. C. Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 67, no. 8, pp. 1558–1572, 2020.